Supporting Information for
Retrieval of Crystallographically-Derived Molecular Geometry Information

Ian J. Bruno, Jason C. Cole, Magnus Kessler, Jie Luo,
W. D. Sam Motherwell, Lucy H. Purkis, Barry R. Smith, Robin Taylor, Richard I.
Cooper, Stephanie E. Harris and A. Guy Orpen

Table S1. CSD reference codes of structures used in testing the bond-type assignment algorithm

A. Structures whose bond types were successfully inferred

AHINUA	AHISMA	AHMALM01	AHMPCU	AHOXLH02	AHTPAL
AHZCCO	AIBAPT10	AICOCU10	AIPBAR	AMPTRA10	AMPYRM10
AMSALA	AMSCNI11	AMSNCO	AMTAZC	AMTETA	AMTETZ01
AMTHZL	AMTRAZ	ATDZSA01	ATHCCO	ATIDIN	ATPGER
ATPONI	ATPRCL10	ATRYMA	ATTDAZ02	ATYREE	ATZACO10
ATZOPI	AUBPHB	BABFIT	BABFUG	BABHER	BABHOB
BABJUJ	BABKEU	BABMEW	BABMUN	BABNEX01	BALREL
BALRUC	BALSEM02	BALSOX	BALTUD	BALVEQ	BALVUF
BALWEQ03	BALWIV	BALWUG	BALXES	BIVZIP01	BIWBEO
BIWCOZ	BIWDEQ	BIWFES	BIWGET10	BIWHIY	BIWJAS01
BIWJUM01	BIWKEX10	BIWLEY	BRTEMN10	BRTPCU	BRUCLB
BRUDAG	BRURID01	BRYNOL10	BRZMSL01	BSBCYO10	BSBZIC01
BSELGY20	BSFULC	BSLTMO	CAHBUI	CAHCEU	CAHCOD
CAHCUJ	CAHDEU	CAHDUK	CAHFEW	CAHFUN	CAHGIB
CAHGUO	CAHHIC	CECXOX	CECYIS	CECZEP	CEDBES10
CEDCET	CEDDUK	CEDFOG	CEDHIC	CEDJIE	CHISCU
CHLRLD	CHLSOS	CHMDIM	CHMHAZ	CHMPIP	CHMPOW01
CHMSDB	CHNPIH10	CHOCHL01	CHOENO	COFNEQ10	COFNUG10
COFPAO10	COFPOC10	COFQIX	COFRIY10	COFSAR	COFSUL10
COFTIA10	COFVEY	COFWEZ	COFXOK	CPFETS	CPFMNI
CPFPMO10	CPFPTI10	CPGXMR20	CPHAZO	CPHCOP	CPHHTZ
CPHMAL10	CPHMUR	CUDDEK	CUDECN	CUDFIQ	CUDGAJ
CUDGOX	CUDHIS	CUDKAN	CUDLAO	CUDMAP	CUDMGC02
CUDMTB20	CUDNIY	DALHED	DALJAB	DALJEF	DALKEG
DALLEH	DALNAF01	DALQAI	DALREN	DALSEO	DALTIT
DEGWAN	DEGXAO	DEHBUN	DEHDID	DEHGIG	DIFERR
DIFFUT	DIFGUU	DIFHIJ	DIFJIL	DIFKEI02	DIFKUY
DIFLUZ	DIFNUB	DIFPOX	DIFQOY	DODNOZ	DODPOB
DODQOC	DODROD	DODSIY	DODTAR	DODVAT	DODWAU
DODXAV	DODYAW	DODZAX	DODZUR	DUNVEN	DUNWEO
DUNXIT	DUNYIU	DUPBEV	DUPCIA10	DUPFEZ01	DUPFOJ10
DUPHEB	DUPJED	DUPKEE	DUPLUV	DUPMOQ	EDEJIG
EDEKIH	EDELIH	EDENAC	EDEPAE	EDEQUZ	EDEWIT
EDEXOA	EDEYUH	EDEZUI	EDIACU20	EFIROA	EFISOB
EFITOC	EFIVOE	EFIXOG	EFIYOH	EFIZOI	EFMXPDP
EFOBOQ01	EFOCUX	EFODUY	ESECSE	ESEXFE10	ESMALP
ESTDOL10	ESTILO01	ESTRON03	ESULYB	ETABPT01	ETACLA
FATQAS	FATREX	FATSAU	FATTUP	FATWAY	FATWOM

FATYOO	FATZOP	FAVBOT	FIWRUY	FIWSUZ	FIWTUA
FIWVUC	FIWXAK	FIWYAL	FIWZAM	FIWZUG	FIXBUJ
FIXCOE	FOXDOF	FIXFOH	FOWCID	FOWDEA10	FOWFEC
FOWGAZ	FOWHAA	FOWLAE	FOWMAF	FOWMUZ01	FOWNUA
FOWPOW	FUJHOH10	FUJJEZ10	FUJJUP10	FUJKIE	FUJLIF
FUJMIG	FUJNED	FUJPEF	FUJQEG	FUJREH	FUJSEI
FUJTEJ	GAKXOF	GAKYEW	GAKZAT01	GALACT	GALBEA
GALCEB	GALDAY	GALDIG	GALGAB	GALHAC10	GALJAE
GEFLIM	GEFMAF	GEFNAG	GEFNOU	GEFQOX	GEFREO02
GEFSEP	GEFTEQ10	GEFVOC10	GETCUD	GETDUE	GETFUG02
GETGOB01	GETHIW	GETJAQ	GETJUK	GETKIZ10	GETLEW
GETMAT	GIWDAR	GIWDUL	GIWFUN	GIWGUO	GIWHOJ
GIWJOL	GIWLED	GIWNEF	GIWPEH	GIWQEI	GOHZIM
GOJBIQ	GOJCIR	GOJDIS	GOJFIU	GOJGIV	GOJHIW
GOJJIY	GOJKEV	GOJLEW	GOJMEX	GOJNEY	GOXFAA
GOXGAB	GOXHAC	GOXJAE	GOXKAF	GOXLAG	GOXNOW
GOXPOY	GOXQEP	GOXREQ	GOXSER	GULQEI	GULREK
GULSEL	GULTOW	GULVOY	GULWOZ	GULXIU	GULZAO
GULZOC01	GUMBOF	GUMCOG	GUMDEX	HABJOJ	HABKAW
HABKIF	HABLAX	HABLEC	HABLUR	HABMEC	HABMON
HABNIH	HABPAB	HABPIJ	HABQAC	HAYREE	HAYSEF
HAYTEG	HAYVOS	HAYWOT	HAYXOU	HAZBAL	HAZBUF
HAZCUG	HEXSOS	HEXTOL	HEXVIP	HEXWOW	HEXYIS
HEXZOS	HEYCIX	HEYDEU	HEYGIB	HODSOI	HODTOJ
HODVOL	HODWOM	HODXUT	HODZAB	HODZUV	HUFJUN
HUFKUO	HUFLUP	HUFNAX	HUFPAZ	HUFQEE	IDELAE
IDEMEJ	IDENEK	IDEPEM	IDEQEN	IGUHOH	IGUJOJ
IGUKOK	IGULOL	IGUPIJ	IGUQIK	IPDXUR01	IPEPNI
IPEYHF10	IPHACN	IPHMGL	IPHTDS	JASDEM	JASFEO
JASHEQ	JASJES	JASKET	JAZBUH	JAZCOC	JAZDIX
JAZFIZ	JAZGIA	JAZHEX	JEPNIB	JEPPID	JEPRED
JEPSEC	JEPTIH	JEPVAB10	JEPWEG	JERRON	JERSEE
JERTAB	JERWAE	JERXEJ	JETHUL	JETJUN01	JETKUO
JIWGIF	JIWHIG	JIWJII	JIWKIJ	JIWLEG	JIWMEH
JOSJUW	JOSKUX	JOSLUY	JOSMOT	JUNPIR	JUNQIS
JUNRIT	JUNTIV	JUNVET	JUPXAT	JUPYAU	KALBEE
KALCAB	KALDAC	KALFIM	KALGOT	KEMPBZ	KEMPUN
KEMQUO	KEMRUP	KEMSOK	KEMTOL	KOJZIS	KOKCES
KOKDET	KOKFEV	KOKGEW	KURTUM	KURVOI	KURWOJ
KURXOK	LATDAL	LATFAN	LATFUH	LATGOC	LATHUJ
LATJUL	LAZYIU	LAZZIV	LCARVX	LCITHC01	LCYLCY10
LCYSTI12	LIRNUV	LIRPUX	LIRSAG	LIRVAJ	LOXLEP
LOXMEQ	LOXNER	LOXPET	LOXQEU	LOXRIZ	LUFQIM
LUFRIN	MAZHEA	MAZHUN	MAZJUS	MAZKON	MAZLOO
MEMANP	MEMBUB	MEMDEN	MEMFEP	MEMGCU	MIZSIX
MIZTIY	MIZVEW	MIZWEX	MIZXEY	MIZYID	MPAMPT10
MPAOTT	MPAPTT	MPASMO	MPATHA	MXOHFE01	MXPACN
MXPCIM	MXPENNA	MXPHIQ01	MXPIMO10	NAMBAE	NAMCAF
NAMDIO	NAMDPC01	NAMFIQ	NARKOG	NARLOH	NARMOI
NARNOJ	NARPOL	NARQOM	NEDKAI	NEDKUC	NEDLUD
NEDMUE	NEDNUF	NEDPUH	NEZJAD	NEZJUX	NEZLOT
NEZMOU	NEZNOV	NIFYUW	NIFZUX	NIGBUA	NIGCUB
NIGDUC	NIGFOY	NIZNAL	NIZPAN	NIZQAO	NIZRAP
NIZSAQ	NIZTAR	NOQXIA	NOQYIB	NOQZIC	NORBAX
NORBUR	NORCET	NURNOD	NURQAS	NURRAT	NURSAU01
NURTUP	ODACIF	ODADIG	ODADUS	ODAFUU	ODAGUV
ODAHUW	OGIMOG	OGIPAV	OGIQAW	OGIROL	OGISOM
OGITON	OPDDTE	OPECIR10	OPENDA	OPHDAH01	OPHPOS

OPIMAS	OTPMOC	OTROSC	OVALDB	OVATAC	OVERAT
PABXOF	PABZAM	PABZEY	PABZSO	PATQOQ	PATROR
PATSOS	PATTOT	PATVEL	PATWEM	PECZAY	PEDBEF
PEDCEG	PEDDCU	PEDERL	PEDFUZ	PEDGIO	PEKLI
PEKNEY	PEKPAW	PEKSIH	PEWPEM	PEWRAK	PEWSAL
PEWTAM	PEWVAO	PICWEA	PICYCO	PICYIJ01	PICZUW10
PIDBUZ	PIXPER	PIXRET	PIXSEU	PIXVEX	POHZAN
POIRCL10	POJBIZ	POJCEW	POJCUM10	POVWOM	POVXUT
POVZEF	POWBEI	POWCEJ	POWDIO	PTCUBR	PTCYPO
PTDETC01	PTECHB	PTERID	PTETHA	PUWNEA	PUWREE
PUWSIJ	PUWTIK	PUWVIM	PUWWIN	PYIPCU	PYMCAM
PYMCXC	PYMECR	PYMNCH10	QANQIF	QANRIG	QANTEE
QANVIK	QANWIL	QANXOS	QEDQAR	QEDQUL	QEDROG
QEZTUK	QEZVUM	QEZXOI01	QEZYQJ	QEZQZK	QIYCEG
QIYDEH	QIYFEJ	QIYGAG	QIYGUA	QIYHUB	QUGYAS01
QUGYUM	QUGZUN	QUHBUQ	QUHHEG01	RAWQOV01	RAWRIQ01
RAWSIR	RAWTAK	RERPIN	RERQIO	RERRIP	RERSIQ
RERTIR	RERVUF01	REYPAM	REYQAN	REYRAO	REYSAP
REYTEU	REYVAS	RIVCEE	RIVFEH	RIVGIM	RIVHIN
ROQWUP	ROQXUQ	ROQYUR	RORCIK	RURQAW	RURRAX
RURSAY	RURTAZ	SAMYAG	SAMYUA	SAMZOV	SANCEP01
SANDEQ	SETRUE	SETSUF	SETVES10	SIJYIT	SIJZIU
SIKBIK	SIKCIY	SIKDIZ	SIKFIB	SOKMAG	SOKMIO10
SOKNIP	SOKPIR	SOKQIS	SOKREP	SUCNEJ	SUCPAC
SUCQIQ	SUKFEJ01	SUKGEK	SUKJEN	SUKJUD01	SUKKUE
SUKMAM	TACRIY	TACRUL	TACSB	TACSNC	TACSUM
TAPQIK	TAPSIM	TAPTIN	TAPVIP	TAWHUU	TAWJOQ01
TAWLAE	TAWMAF	TAWNAG	TAWPAI	TEJMAW01	TEJNAX
TEJPAZ	TEJQEE	TEJREF	TEJSEG	THACPU23	THALID03
THAMCU01	THANPT	THAZPN	TIMSUD	TIMTOY	TIMVOA
TIMWOB	TIMXOC	TMAHGB01	TMAMAL	TMAMCD05	TMAMCD14
TMAMMC	TMAMNB	TOXISA10	TOXJOF	TOXLAT	TOXMOI
TOXNID	TOXNUP	TUYNQ	TUYPOS	TUYQOT	TUYROU
TUYSOV	TUYTOW	UFOLAC	UFONAE	UFOPAG	UFOQAH
UFORAI	UHEJIA	UHEKIB	UHELIC	UHEMID	UHENIE
UHEPIG	VAGYEI	VAGYUX	VAGZIM	VAHBIP	VAHCIQ
VAHDEN	VAVJOR	VAVLEJ	VAVMEK	VAVNEL	VAVPIR
VECCOV	VECDUC	VECGAL	VECGUF10	VECHOA	VECKAP
VEQWOD	VEQXOE	VEQYOF	VERBEZ	VERCEA	VIFHAT
VIFJAV	VIFKAW	VIFLAX	VIFMAY	VIFNAZ	VITDAD
VITGIO	VITHIP	VITJEN	VORJIT	VORJIV	VORKAO10
VORMEU	WABNIW	WABPIY	WABPUL	WABRAS	WAVTAO
WAVVAQ	WAVWAR	WAVXAS	WAVYAT	WERPAK	WERQIT
WERRIU	WERSIV	WERTIW	WERVIY	WIGHUP	WIGKAY
WIGLAZ	WIGMAA	WIGPAD	WINCUR01	WINDUS	WINFOO
WINGOP	WINHUW	WOSCEM	WOSDEN	WOSFEP	WOSGEQ
WOSHER	WOSJET	WURCAN	WURDES	WURFEU	WURGIZ
WURHIA	XAXCUU	XAXDOP	XAXFOR01	XAXGAE	XAXHAF
XIDMOM	XIDNON	XIDPOP	XIDRIL	XIDTEJ	XOKVIC
XOKWID	XOKXUQ	XUKFAK	XUKGAL	XUKHAM	XUKJAO
XUKJIW	XUZNUB	XUZPUD	XUZRIT	XUXTER	XUZVET
YAPYIX	YAPZEU01	YAQBIB	YAQCIC	YAQDID	YAQLAD01
YEMZIZ	YENBIC	YENBOI	YENCID	YENDEA	YENFAY
YICKIE	YICLIF	YICMIG	YICNIH	YICQIK	YISKAM
YISKUG	YISLUH	YISMOC	YISNOD	YOJFAE	YOJGAF
YOJHEK	YOJJIQ	YOJKUD	YOJMIT	YOZVAK	YOZKOA
YTANPY10	YUBBUS	ZABPOH	ZABRAV	ZABSEA	ZABTOL
ZABXAB	ZAYNUI	ZAYPUK	ZAYQUL	ZAYRUM	ZAYTAU

ZAYVAW	ZEGMIH	ZEGNII	ZEGNOO10	ZEGQAD	ZEGREI
ZEWDAG	ZEWFOW	ZEWGUD	ZEWJUG	ZILPUF	ZILQUG
ZILRUH	ZILSUI	ZILTUJ	ZILVUL	ZNETAM10	ZNGLUD
ZNKAIH	ZNOMEX	ZNPICH01	ZOXXAL	ZOXYUG	ZOYBEU
ZOYDIA	ZZZBEM01	ZZZBFM02	ZZZLAO10	ZZZTKA01	ZZZTLC01

B. Structures whose bond types were not successfully inferred

AHPAZP10	AHPNMC	AMPURM	AMPYMO	ATNNIA	BABGOA
BABKUK	BALTEO	BIWCEP	BSALAN	CAHBIX	CEDGOH
CEDCUJ	CEDKIF	CHOEST20	CPFYBA10	CPHECR	CUDJEQ
DALPAH	DEGTIS	DEGVEQ	DEGYAP	DEGYUJ	DEGZUK
DEHCOI	DEHFIF	DIFMUA	EDEQAF	ESTRON12	FATTAV
FATXIH	FAVCIO	FOWJAC	FOWKAD	GALFAA	GEFKIL
GEFVAO	GETFEQ02	GETMPH	GIWKOM	GIWMEE	GOXMOV
HAYYIP	HAYZEM	HAZDOB	HEXXOX	HEYBOC	HEYFIA
IDERAK	IGURIL	JASGEP	JERTUV	JETFIX	JETGOE
JOSHAA	JOSHUU10	JUNSIU	KAKZIF	KOKBIV	KURSUL
KURZEC	LIRQUY	LIRTAH	LUFLAZ	LUFMEE	LUFNEF
LUFPILO1	MAZGEZ	MELZEI	MEMCOW	MPAZPC	MXPBUQ
NAMGOX	NEZKOS01	NURSOI	OTPHMO	PABYAT	PABYIB
PEKQEB	PEKRIG	PEWRUE	PICXEE	PIXQAO	PIXTEV
POJDOH	PYLTCF	QEDMAN	QEDNAO	QEDPAQ	QEZWUN
QUHCUR	RAWQAH	RAWVAM	RIVBED	RIVDEF	ROQZOM
RORBIJ	RURNAT	RURPAV	SANBEO	SETTMO10	SETWET
SETXIY	SUCMOS	SUCPOV	SUCREN	SUKHEL	TACRFE10
TAPRIL	TAPUFE	THBRNB10	TIMYIX	UFOMAD	VAVKIM
VEQZOG	VITAMB	VITCAC	VORLAP	VORNAR20	WABNUJ
WABQIA	WAVSAN	WECSOM	WECTON	WECVOP	WECWUW
WECXUX	WECYOS	WIGNAB	WINCED	WURBAM	XAXBON
XIDQOQ	XOKRIY	XOKSIZ	XOKTIA	XUKDAI	XUZWEU
YICPIJ	YISQAS	YOZWAL	YOZYUH	ZABVUT	ZEGPAC
ZEWDUA	ZEWHUE	ZNPORT01	ZOXXUF01	ZOYCEV	ZZZTMK01

Table S2. Results from exact searches on bond fragments

CSD entry	bond	N	$\sigma(\text{\AA})$	$\Delta(\text{\AA})$
AHONAM	C1-C2	2313	0.016	0.012
AHONAM	C1-C6	62	0.010	0.041
AHONAM	C5-C6	4367	0.013	0.030
AHONAM	C6-C7	36	0.014	0.028
AHONAM	C8-C13	4896	0.016	0.009
AHONAM	C9-C10	4801	0.019	0.031
AHONAM	C10-C11	4832	0.026	0.085
AHONAM	C26-C27	1620	0.069	0.028
AHONAM	P1-C14	301	0.013	0.045
AHONAM	O3-C27	842	0.064	0.112
AHOPIW	C1-C4	105	0.020	0.008
AHOPIW	C2-C14	17	0.008	0.011
AHOPIW	C10-C11	8	0.033	0.002
AHOPIW	C14-C15	4199	0.014	0.003
AHOPIW	C20-C21	4864	0.015	0.005
AHOPIW	N1-C10	19	0.013	0.016
AHOPIW	O1-C11	1795	0.027	0.008
AHOPIW	P1-C20	16	0.009	0.005
AHOPIW	S1-C3	6	0.028	0.040
BASKAI	C10-C11	50	0.009	0.013
BASKAI	C12-C13	3261	0.027	0.024
BASKAI	C16-C17	195	0.018	0.006
BASKAI	C16-C18	553	0.026	0.001
BASKAI	C17-C20	431	0.026	0.006
BASKAI	C22-C23	9	0.010	0.015
BASKAI	N1-C1	12	0.009	0.000
BASKAI	N1-O2	157	0.019	0.003

BASKAI	N2-C1	18	0.011	0.003
BASKAI	N2-O2	171	0.018	0.004
BASKAI	N3-C10	692	0.015	0.017
BASKAI	N4-C11	6	0.012	0.007
BASKAI	N4-C12	2631	0.018	0.006
BASKAI	N5-C16	133	0.016	0.001
BASKAI	N6-C17	167	0.021	0.003
BASKAI	N7-C22	11	0.006	0.004

Table S3. Results from generalised searches on bond fragments

CSD entry	bond	N	$\sigma(\text{\AA})$	$\Delta(\text{\AA})$
IHIYED	B1-F1	17	0.043	0.059
IHOGUH	C1-C6	18	0.014	0.010
IHOGUH	C5-C6	15	0.014	0.004
IHOHUI	C25-C26	153	0.008	0.004
IHOJOE	C7-C15	15	0.018	0.006
IHOJOE	C15-C16	21	0.024	0.006
IHOQUR	C10-C14	17	0.015	0.002
JAFKOR	O2-C5	15	0.011	0.019
JAFLEI	O3-C2	17	0.011	0.023
JAFMOT	P1-C9	30	0.011	0.009
JAFMOT	S1-C1	21	0.008	0.001
JAFNOU	P2-C1	15	0.011	0.014
JAFQEN	C1-C6	81	0.016	0.015
JAFREO	S1-C2	15	0.005	0.021
JAFSOZ	N1-N2	7	0.017	0.065
JAFSOZ	P1-C6	83	0.014	0.019
JAFVUI	O1-C1	60	0.046	0.008
JAFXEU	C12-C13	20	0.009	0.014
KABKAA	S1-N1	14	0.036	0.026
KABLIJ	Si17-Si19	16	0.021	0.027

Table S4. Comparison of exact and generalised searches on bond fragments

CSD entry	bond	<u>exact</u>			<u>generalised</u>		
		N	$\sigma(\text{\AA})$	$\Delta(\text{\AA})$	N	$\sigma(\text{\AA})$	$\Delta(\text{\AA})$
AHOPIW	C2-C14	17	0.008	0.011	83	0.018	0.013
AHOPIW	P1-C20	16	0.009	0.005	75	0.009	0.007
AHOREU	C18-C19	19	0.010	0.004	65	0.034	0.011
AHOREU	C18-C20	13	0.010	0.014	42	0.010	0.012
AHORIY	C6-C7	22	0.012	0.001	53	0.013	0.004
BASKAI	N1-C1	12	0.009	0.000	112	0.026	0.004
BASKAI	N6-C15	18	0.011	0.008	150	0.019	0.017
BASPAN	N1-C1	24	0.015	0.005	50	0.014	0.001
BASPAN	N1-C9	18	0.008	0.007	36	0.013	0.006
BASWOI	C15-C16	12	0.010	0.016	53	0.066	0.028
BASWOI	N1-C7	15	0.017	0.012	56	0.013	0.020
BASWUO	C1-C8	18	0.014	0.011	36	0.016	0.007
BASWUO	O3-C27	26	0.017	0.018	103	0.022	0.016
BASXEZ	N2-C30	12	0.007	0.002	24	0.020	0.002
BASXEZ	O4-C35	16	0.014	0.014	34	0.013	0.011
BASXID	Br2-C44	24	0.018	0.002	173	0.026	0.000
BASZUR	C1-C2	17	0.008	0.021	40	0.012	0.028
BASZUR	C9-C10	16	0.016	0.006	34	0.019	0.011
BATCUV	B1-C13	15	0.005	0.006	244	0.014	0.008
EKEMIQ	N1-C1	19	0.011	0.006	81	0.012	0.014
EKEMIQ	O1-C7	11	0.009	0.012	22	0.012	0.009
EKERIV	C15-C19	12	0.016	0.007	86	0.016	0.012
EKERUH	N1-C4	10	0.009	0.000	48	0.033	0.002
EKEYEY	N1-C2	12	0.009	0.010	59	0.009	0.010
EKEYOI	N3-C4	14	0.048	0.025	35	0.032	0.023
EKEYUO	N2-C18	19	0.009	0.007	43	0.013	0.002
EKEZEZ	N2-C13	12	0.010	0.002	41	0.010	0.005
EKEZID	C2-C3	20	0.018	0.000	60	0.018	0.000

EKIDUX	O2-C16	19	0.035	0.008	51	0.032	0.012
EKIPET	C5-C11	10	0.020	0.005	22	0.017	0.010
EKODEN	N1-C2	11	0.021	0.008	34	0.014	0.013
EKODEN	N1-C4	35	0.014	0.024	73	0.017	0.024
EKOKAQ	C13-C19	10	0.011	0.017	20	0.018	0.004
EKOXUX	P1-C11	10	0.016	0.009	23	0.017	0.004
EKUHOH	O1-C10	13	0.016	0.004	106	0.017	0.006
EKUJOJ	O4-C6	14	0.014	0.007	32	0.014	0.008
ELEJOU	C6-C7	15	0.012	0.005	376	0.029	0.012
ELEMEN	P1-C8	16	0.009	0.003	75	0.009	0.005
IHIWUR	P1-O1	9	0.029	0.002	36	0.017	0.001
IHOGUH	C1-C7	21	0.013	0.002	96	0.015	0.001
IHOHUI	C17-C22	16	0.013	0.003	73	0.014	0.000
IHOJOE	C8-C9	32	0.009	0.002	64	0.009	0.004
IHOQUR	C3-C8	20	0.017	0.013	44	0.017	0.016
IHOQUR	C9-C10	12	0.009	0.032	26	0.014	0.021
JAFKOR	O1-C1	21	0.065	0.008	48	0.047	0.004
JAFMOT	C9-C10	10	0.016	0.007	123	0.019	0.019
JAFNOU	P1-O3	12	0.009	0.007	32	0.009	0.005
JAFREO	B1-C1	30	0.011	0.013	112	0.015	0.012
JAFREO	C2-C5	18	0.061	0.038	61	0.045	0.005
JAFVUI	P1-C6	16	0.006	0.007	199	0.016	0.016

Table S5. Results from exact searches on valence angle fragments

CSD entry	valence angle	N	$\sigma(^{\circ})$	$\Delta(^{\circ})$
AHONAM	C1-C6-C5	52	0.7	0.2
AHONAM	C1-C2-C3	1982	2.0	0.7
AHONAM	C2-C1-C6	55	0.9	1.8
AHONAM	C4-C5-C6	4311	1.4	0.5
AHONAM	C5-C6-C7	22	2.1	0.5
AHONAM	C8-C9-C10	4935	1.2	1.1
AHONAM	C9-C10-C11	4746	1.2	0.7
AHONAM	C10-C11-C12	4712	1.4	2.7
AHONAM	C14-P1-C20	216	2.1	3.8
AHONAM	C15-C14-C19	4924	1.4	0.8
AHONAM	C26-C27-C28	842	6.3	3.2
AHONAM	P1-C14-C15	602	2.0	1.9
AHONAM	O3-C27-C26	1620	6.4	8.3
AHOPIW	C1-C4-C5	210	1.5	0.6
AHOPIW	C2-C14-C15	34	1.0	1.6
AHOPIW	C10-N1-C13	7	3.3	3.1
AHOPIW	C11-O1-C12	901	2.5	0.6
AHOPIW	C14-C15-C16	4432	1.1	0.4
AHOPIW	C15-C14-C19	4150	1.2	0.7
AHOPIW	C20-C21-C22	4948	1.2	0.3
AHOPIW	C20-P1-C26	8	1.5	0.8
AHOPIW	C21-C20-C25	4879	1.2	0.6
AHOPIW	N1-C10-C11	6	1.1	2.0
AHOPIW	O1-C11-C10	8	1.4	0.1
AHOPIW	P1-C20-C21	32	1.7	1.6
AHOPIW	S2-P1-C20	16	1.8	1.2
BASKAI	C8-N3-C10	11	0.7	0.5

BASKAI	C11-C10-C14	40	0.6	1.2
BASKAI	C12-C13-C14	501	1.1	1.7
BASKAI	C15-N5-C16	12	1.5	0.9
BASKAI	C15-N6-C17	18	0.9	0.8
BASKAI	C16-C17-C20	387	2.3	0.8
BASKAI	C17-C16-C18	386	2.2	0.4
BASKAI	C18-C16-C19	275	3.1	0.4
BASKAI	C20-C17-C21	208	3.2	0.1
BASKAI	N1-C1-N2	9	1.1	0.5
BASKAI	N3-C10-C11	38	0.9	2.4
BASKAI	N3-C10-C14	652	2.0	0.3
BASKAI	N4-C12-C13	2222	1.4	1.6
BASKAI	N5-C16-C17	119	1.9	0.5
BASKAI	N5-C16-C18	265	2.5	2.3
BASKAI	N6-C17-C16	160	1.5	0.5
BASKAI	N6-C17-C20	333	2.5	1.9
BASKAI	N7-C22-C23	9	0.7	0.5
BASKAI	O1-N1-C1	12	1.1	0.5
BASKAI	O1-N1-C2	133	1.7	0.2
BASKAI	O2-N2-C1	18	1.0	0.1
BASKAI	O2-N2-C3	167	2.0	0.5

Table S6. Results from generalised searches on valence angle fragments

CSD entry	valence angle	<i>N</i>	$\sigma(^{\circ})$	$\Delta(^{\circ})$
AHONAM	N1-C1-C2	65	1.7	1.6
AHOPES	C4-N1-C25	18	1.5	1.1
AHOPIW	C2-P1-C20	12	2.1	1.3
AHORAQ	C2-C1-C22	17	2.5	2.2
AHOREU	C14-C15-C17	15	2.5	1.2
AHORIY	C4-C5-C6	15	1.5	4.0
AHOTIA	C13-C10-C17	19	4.3	16.2
BASKAI	N3-C8-C1	16	2.8	1.2
BASPAN	C1-N1-C9	15	0.8	0.9
BASWOI	C1-C2-C3	18	2.8	1.1
BASWUO	C1-N1-C2	34	0.4	0.1
BASXEZ	S1-O4-C35	9	1.9	3.0
BASXID	Br2-C44-C45	22	2.1	0.7
BASZUR	C4-C5-C10	15	2.7	4.8
BATCUV	C1-B1-C25	9	4.3	8.6
EKEMIQ	C1-C6-C7	15	3.3	1.9
EKEYOI	C3-N3-C4	16	2.0	3.7
EKEYUO	C11-C10-C19	15	2.9	0.1
EKEZEZ	C3-N1-C4	16	2.7	0.3
EKEZID	S1-C5-C1	15	1.8	2.1

Table S7. Comparison of exact and generalised searches on valence angle fragments

CSD entry	valence angle	<u>exact</u>			<u>generalised</u>		
		<i>N</i>	$\sigma(^{\circ})$	$\Delta(^{\circ})$	<i>N</i>	$\sigma(^{\circ})$	$\Delta(^{\circ})$
AHONAM	C5-C6-C7	22	2.1	0.5	82	2.2	1.7
AHOPIW	O1-C11-C10	8	1.4	0.1	24	1.5	0.1
AHOPIW	S2-P1-C26	16	1.8	1.6	71	1.8	1.1
AHOREU	C36-C41-C40	19	0.6	0.5	76	0.9	0.9
AHOREU	C44-C43-C45	13	1.6	2.7	101	2.6	1.9
AHORIY	C1-O4-C4	23	2.1	3.2	188	2.8	2.7
AHOTIA	C15-C14-C19	10	0.5	0.2	25	2.2	0.2
BASKAI	C1-N1-C2	12	1.5	0.5	111	1.9	0.1
BASKAI	O4-N6-C15	18	1.0	0.3	150	1.7	0.4
BASPAN	N1-C1-C6	24	2.0	1.1	50	1.7	0.4
BASPAN	O1-C11-O2	9	3.5	2.3	20	3.2	4.8
BASWUO	N2-C12-O2	17	1.9	0.0	55	1.7	0.9
BASXEZ	N2-C30-C31	11	1.4	2.4	24	1.7	2.1
BASXID	N4-C52-O8	18	1.0	0.1	44	1.5	0.5
BASZUR	C2-C3-C4	12	1.2	1.0	44	1.3	1.3
BASZUR	C20-C19-C24	16	1.1	0.9	38	1.2	1.0
BATCUV	C19-B1-C25	15	4.5	6.0	358	5.7	5.3
EKEMIQ	N1-O1-C7	11	0.7	1.8	22	1.1	2.0
EKERIV	O1-C11-O2	11	8.5	1.6	23	5.8	1.8
EKEYEY	N1-C2-C20	12	0.4	0.4	39	2.5	0.2
EKEYOI	O1-C4-N3	14	2.4	3.6	34	1.8	3.1
EKEZEZ	S2-N2-C13	12	2.0	2.7	54	2.1	1.2
EKEZID	C2-N1-C4	13	2.5	6.5	27	2.6	5.8
EKEZID	N2-C3-C2	20	1.0	0.9	53	2.1	0.6
EKIDUX	N1-C1-C2	15	1.7	1.8	50	2.9	1.7
EKIDUX	O2-C16-C15	19	1.2	0.1	42	2.3	1.1
EKIPET	C11-C10-C12	10	1.7	1.8	51	2.3	1.4
EKODEN	C2-N1-C4	10	1.1	1.1	32	1.9	2.0

EKODEN	N1-C2-C3	11	1.2	0.2	34	1.2	0.1
EKOKAQ	C14-C13-C19	20	1.0	1.5	40	1.1	1.6
EKOKEU	N1-C8-C9	10	2.3	1.7	20	2.0	1.2
EKOXUX	P2-C19-C21	30	3.5	2.6	63	4.1	2.9
EKOYUY	N3-C13-C14	14	2.4	2.5	62	2.3	2.5
EKUHOH	C1-C2-C7	12	1.1	0.4	24	1.7	0.7
EKUHOH	O1-C10-C11	11	0.2	0.2	89	0.7	0.7
EKUJOJ	C4-C3-C5	30	1.0	0.0	65	1.3	0.2
EKUJOJ	O4-C6-C1	10	1.5	1.3	39	1.8	1.9
ELELEM	C7-P1-C13	13	1.9	3.9	26	2.3	3.0
ELEMEN	N2-C2-C3	11	1.2	0.8	160	1.1	0.4
ELEMEN	S1-P1-C14	16	1.8	1.4	71	1.8	1.9
IHIVUQ	C16-C17-C18	28	1.3	2.9	66	1.3	3.4
IHOFEQ	C34-C35-C36	30	1.1	0.7	4477	1.1	2.1
IHOGUH	N1-C7-C1	21	1.6	0.9	94	1.7	0.7
IHOHUI	Br1-C1-C6	16	1.0	0.9	65	2.1	1.3
IHOHUI	C23-N2-C25	15	0.7	0.6	41	1.1	0.6
IHOQUR	C6-C7-C8	13	2.5	3.0	55	2.5	5.1
IHOQUR	O1-C11-C10	28	1.3	1.7	97	1.3	2.1
JAFKOR	N1-C1-O1	21	11.2	2.5	47	7.7	1.4
JAFNOU	P1-O2-C3	12	2.3	1.9	24	2.3	2.7
JAFREO	C2-C3-C6	18	2.0	0.7	44	1.8	0.7

Table S8. Results of searches on torsion fragments

CSD entry	torsion	<i>N</i>	% bars occupied	query in occupied bar?	query in local peak?	$d_{min}(^{\circ})$
AHOPES	C2-C1-C7-C12	483	50.0	yes	no	0.1
AHOPES	C3-C2-C1-C7	32	35.0	yes	no	0.1
AHOPES	C28-N1-C4-C29	22	37.5	yes	no	0.8
AHOPIW	N1-C1-C4-C5	32	37.5	yes	yes	1.0
AHOREU	C42-N2-C43-C45	19	35.0	no	no	0.3
AHOREU	N2-C42-C40-C41	656	47.5	yes	no	0.2
AHORIY	C7-C6-C5-C11	31	30.0	yes	yes	0.7
BASWOI	C8-C1-C9-C10	30	47.5	no	no	6.9
BASWOI	C8-C1-C15-C16	30	17.5	yes	no	0.1
BASWOI	O2-C16-C15-C1	31	27.5	yes	yes	0.2
BASWUO	C12-N2-C20-C21	98	30.0	yes	no	1.0
BASWUO	O3-C27-C19-C11	28	40.0	no	no	2.9
BASXEZ	C35-O4-S1-C36	26	32.5	yes	no	0.1
BASXEZ	F3-C36-S1-O4	84	37.5	yes	no	0.1
BASXID	C36-N3-C37-C38	116	40.0	no	no	3.8
BASXID	O3-C24-N2-C17	39	32.5	yes	yes	0.1
EKEMIQ	C8-C13-O4-C15	2004	50.0	yes	no	0.0
EKEMIQ	O1-C7-C8-C9	23	37.5	yes	yes	2.1
EKEYEY	C21-N6-C23-C28	44	50.0	no	no	3.2
EKEYEY	C21-N7-N8-C29	30	37.5	no	no	2.9
EKEYOI	C3-N3-C5-C6	46	22.5	yes	no	0.1
EKEYOI	N1-C3-C2-C1	34	30.0	yes	yes	0.2
EKEYUO	C14-C16-C32-C37	56	45.0	yes	no	1.0
EKEYUO	C16-C17-C38-C43	42	37.5	yes	yes	0.3
EKEZEZ	C6-S2-N2-C13	72	22.5	yes	no	0.0
EKIDUX	C15-C16-O3-C17	37	10.0	yes	yes	0.2
EKIFAF	C6-C7-N1-C8	30	40.0	yes	no	1.5

EKINAN	C1-N1-C10-C15	43	42.5	yes	yes	0.1
EKIREV	O2-S1-S2-O4	130	32.5	yes	no	1.3
EKOCOW	N2-C19-N3-C28	31	25.0	yes	yes	0.1
EKOCOW	O1-C19-N2-C20	64	20.0	no	no	3.1
EKODEN	C1-N1-C4-C5	39	32.5	yes	no	0.3
EKOKAQ	P1-C19-C13-C18	26	40.0	yes	no	2.2
EKOKEU	C6-C7-N1-C8	32	35.0	yes	yes	0.3
EKOKEU	O3-C11-N2-C10	21	27.5	yes	yes	0.1
EKOXUX	S1-P1-C11-C12	30	37.5	yes	no	0.6
EKOXUX	Si1-N3-Si2-C28	42	42.5	yes	yes	1.2
EKUHOH	O1-C10-C9-C8	21	15.0	no	no	2.4
ELEJOU	C7-N3-C8-C9	256	40.0	yes	yes	0.1
ELEMEN	N2-N1-C1-N3	35	42.5	yes	no	0.6
IHITOI	B1-N2-C1-C3	56	50.0	yes	no	0.6
IHITOI	C1-N2-B1-C8	32	45.0	yes	no	2.2
IHIWUR	S4-C4-N2-C6	126	12.5	yes	yes	0.0
IHOGUH	C1-C6-C15-C16	58	47.5	yes	no	0.4
IHOGUH	O1-C8-C1-C6	46	45.0	yes	yes	0.4
IHOHUI	N2-C25-C26-C27	32	30.0	yes	yes	0.5
IHOHUI	O3-C23-C22-C17	36	25.0	no	no	5.9
IHOJOE	C19-C20-C21-C22	41	32.5	yes	no	0.6
IHOJOE	C21-C22-C23-C24	70	45.0	yes	yes	0.0
IHOQUR	C8-C9-O2-C18	30	40.0	yes	yes	0.1